# CERTIFICATE

I, the undersigned, Ken SATO, residing at 2-8-12, Koenji-kita, Suginami-ku, Tokyo, JAPAN, hereby certify that to the best of my knowledge and belief the following is a true translation into English made by me of Japanese Patent Application No. 1999-185859 filed on June 30, 1999.

Date: September 24, 2003

K. Sato

Name: Ken Sato

[Document] Patent Application

[Filing Number] A009902405

[Filing Date] 30th June, 1999

[Addressed to] The Commissioner of the Patent Office

[International Patent Classification] G10L 3/00

[Title of the Invention] Voice recognition supporting method and voice recognition system

[Number of Claims] 9

[Inventor]

[Address or Residence] c/o K.K. Toshiba Kansai Research Center, 8-chome 6-26, Motoyamamami-cho Higashinada-Ku, Kobe-shi, Hyogo-ken

[Name] Mitsuyoshi TACHIMORI

[Inventor]

[Address or Residence] c/o K.K. Toshiba Kansai Research Center, 8-chome 6-26, Motoyamaminami-Cho, Higashinada-ku, Kobe-shi, Hyogo-ken

[Name] Hiroshi KANAZAWA

[Applicant]

[Identification Number] 000003078

[Name] Kabushiki Kaisha Toshiba

[Agent]

[Identification Number] 100058479

[Attorney]

[Name] Takehiko SUZUE

[Phone Number] 03-3502-3181

[Appointed Sub-Agent]

[Identification Number] 100084618

[Attorney]

[Name] Sadao MURAMATSU

[Appointed Sub-Agent]

[Identification Number] 100068814

[Attorney]

[Name] Atsushi TSUBOI

[Appointed Sub-Agent]

[Identification Number] 100092196

[Attorney]

[Name] Yoshiro HASHIMOTO

[Appointed Sub-Agent]

[Identification Number] 100091351

[Attorney]

[Name] Satoshi K0NO

[Appointed Sub-Agent]

[Identification Number] 100088683

[Attorney]

[Name] Makoto NAKAMURA

[Appointed Sub-Agent]

[Identification Number] 100070427

[Attorney]

[Name] Shoji KAWAI

[Indication of Fees]

[Pre-payment Ledger Number] 011567

[Amount Paid] 21,000 Yen

[List of Items filed]

[Name of Item] Specification 1

[Name of Item] Drawing 1

[Name of Item] Abstract 1

[Requirement of Proof] Required

[Document] Specification

[Title of the Invention] Voice recognition supporting method and voice recognition system

[Claims]

[Claim 1]

A voice recognition supporting method characterized in that it is applied to a system capable of geographical searching by a voice,

acquires the recognized result by recognizing input user voice,

judges a distance from a point representing the recognized result to a reference point that is set as a reference position of the geographical searching when the recognized result represents a point on a map, and

when the point representing by the recognized result is judged far from the reference point, generates a confirming response for urging user to confirm the right or wrong of the recognized result and presents the confirming response to user.

[Claim 2]

A voice recognition support method characterized in that it is applicable to a system capable of vocally geographical searching,

acquires plural recognition candidates by recognizing input voice of a user as a recognized result, and when the first rank candidate of the recognized result represents

a point on a map,

extracts a recognized candidate representing a point on a map from the recognized result when the first rank candidate in the recognized result represents a point on a map,

executes the re-scoring of collated scores representing similarity with input voice or a distance of each of the extracted recognized candidates to convert into new scores added with a distance between the point represented in the said recognized result and a reference point that is set as the reference position of the geographical searching,

judges a distance from a point represented by the first rank candidate that is decided by the new score of each recognized candidate after the re-scoring to the said reference point, and

when the point represented by the first rank candidate after the re-scoring is judged far from the reference point, generates a confirming response for urging a user to confirm right or wrong of the recognized result for the higher rank candidates up to prescribed numbers after the re-scoring and presents that confirming response to a user.

[Claim 3]

A voice recognition supporting method according to Claim 1 or 2, characterized in that the judging standard for judging a distance from the point represented in the

recognized result to the reference point is changed and set according to prescribed parameters.

[Claim 4]

A voice recognition supporting method according to Claim 1 or 2, characterized in that the standard range for judging a distance including the reference point is varied and set according to prescribed parameters and a distance from a point represented by a recognized result to the reference point is judged from the relation of position of the said point to the reference range.

[Claim 5]

A voice recognition supporting method according to Claim 1 or 2, characterized in that reliability of the recognized result that is subject to the distance judgment is judged and only when judged not reliable, the distance judgment is executed.

[Claim 6]

A voice recognition supporting method according to Claim 1 or 2, characterized in that words representing points on a map subject to recognition are layered according to prescribed segments on the map and controlled, and

When the recognized result represents a point on the map, a segment of the layer that becomes the reference for the judgment of a distance is decided from that point and the distance is judged according to that segment itself or the relation of that segment with a segment to which

the reference point belongs.

[Claim 7]

A voice recognition supporting method according to Claim 1 or 2, characterized in that an attribute that designates the reference for the distance judgment is assigned in advance to each recognizing vocabulary representing a point on the map and

a distance to the point representing the recognition result is judged according to the judging standard designated by the attribute of the recognition vocabulary of the recognized result.

[Claim 8]

A voice recognition system characterized in that it is a voice recognition system that is applied to a system capable of the geographical searching by a voice, comprising:

a reference point setting means to set up reference points that become reference positions in the geographical searching,

a voice recognition means to recognize a voice input by a user and acquire its recognized result,

a distance judging means to judge a distance from a point representing by a recognized result to a reference point when the recognized result acquired by the voice recognition mean represents a point on the map,

a response generating means to generate a confirming

response for urging a user to confirm right or wrong of a recognized result when the point representing the recognized result is judged to be far from the reference point by the distance judging means, and

a presenting means to present the recognition response generated by the response generating means to a user.

[Claim 9]

A voice recognition system characterized in that it is a voice recognition system that is applied to a system capable of the geographical searching by a voice, comprising:

a reference point setting means to set up reference points that become reference positions in the geographical searching,

a voice recognizing means to recognize a voice input by a user and acquire plural recognized candidates as recognized results,

when the first rank candidate in the recognized result acquired by the voice recognizing means represents a point on a map, extracts a recognizing candidate representing a point on the map from the recognized result and

a re-scoring means to re-score a collated score representing similarity with or a distance to each input voice of the recognized candidate representing a point on the map extracted into a new score added with a distance between the point representing the recognized result and

the reference point,

a distance judging means to judge a distance from the point represented by the first rank candidate that is decided by a new score of each recognizing candidate after the re-scoring by the re-scoring means to the reference point,

a response generating means to generate a confirming response for urging a user to confirm right or wrong of the recognized result for higher rank candidates up to prescribed numbers after the re-scoring when the point represented by the first rank candidate after the re-scoring by the distance judging means, and

a presenting means to present a confirming response generated by the response generating means to a user.

[Detailed Description of the Invention]

[0001]

[Technical Field of the Invention]

The present invention relates to a voice recognition supporting method suited to a system capable of geographical searching by voice and voice recognition system.

[0002]

[Prior Art]

In a system capable of the geographical searching by voice that is resented by a car navigation system, a user speaks a place or a name of a facility (a place name or

facility name) in a searching zone in many cases. Accordingly, it is highly possible that a point represented by a recognition result is a place far away from a searching area or a facility that is erroneously recognized. Therefore, when a system operates unconditionally for such the recognition result (for instance, the operation to expand and display a map near the place representing the recognition result, it results in an erroneous operation in many cases.

[0003]

So, in the case of a currently available car navigation system, when the geographical searching is made by a voice command, plural areas are made in the hierarchy structure and an area into which a voice command is input is restricted, and an area that is subject to the search (or a vocabulary specifying it) is more restricted whenever moving from a higher layer down to a lower layer. However, in order to input a name of place outside the limited range in the state wherein an area into which a voice command is input, it is necessary to once release the restriction and thus, a system becomes worse to use.

[0004]

[Problems to be solved by the Invention]

Thus, in a conventional geographical searching system represented by a car navigation system capable of geographical searching by a voice command, there was such

a problem that when an area into which a voice command can be input is restricted by making areas in the layered state; that is, a place (or a vocabulary presenting it into which a voice can be input is restricted, it was necessitated to once cancel the restriction and thus, usability of the system becomes worse.

[0005]

The present invention is made in view of the above circumstances and its purpose is to provide a voice recognition supporting method and a voice recognition system capable of improving usability of a system by avoiding unnecessary erroneous operations without restricting the range of vocabulary indicating places, etc. on a map that can be vocally specified.

[0006]

[Means for solving the problems]

The voice recognition support method of the present invention is characterized in that when a voice recognition result for a voice command input from a user indicates a point on a map, a distance form that point indicating the recognition result to a reference point that becomes a reference position in the geographical searching, that is, whether the point indication by the recognition result is far from or close to the point represented by the recognition result is judged and when judged far, that is, the point represented by the recognition result is a place far from

the reference point, a confirming response (a confirming message) is presented to a user for urging the confirmation of right or wrong of the recognition result and ask a user's direction. Here, it is better to use a current position of a system (or a substance equipped with a system, for instance, a vehicle or a user carrying a system) or an aimed point, etc. This reference point can be changed and set corresponding to a user request or by the autonomic action of a system.

[0007]

Thus, in this invention, when a place or a facility represented by the voice recognition result is judged far from such a reference point as a current location or an aimed point, a confirming response is presented to a user urging to confirm the recognition result for obtaining a user's direction and it is therefore possible to avoid unnecessary erroneous actions by erroneous recognition and improve usability of the system without restricting a range of vocabulary indicting places, etc. on a map that can be specified vocally.

[0008]

Thus, in the case of a system wherein plural recognized candidates are obtained as a voice recognition result, when the first rank candidate of the recognition result represents a point on a map, it is possible to extract recognized candidates on a map from the recognition results,

re-score a collated score (an evaluated value representing similarity or a distance) with input voices of the extracted recognized candidates to a new score added with a distance between the point representing the recognition result and the reference point and judge a distance represented by the first ranked candidate that is decided by a new score of each recognition candidate after the re-scoring, and when a point presented by the first ranked candidate is judged far, it is better to generate a confirming response and present it to a user urging the user to confirm right or wrong of the recognition result for higher candidates limited to the specified numbers after the re-scoring. In this voice recognition support method, it is possible to avoid unnecessary erroneous actions by erroneous recognition and improve usability of the system without restricting vocabulary (a range subject to searching).

[0009]

Further, the present invention features that a reference for judging a distance is changed and set according to a prescribed parameter. When a distance from a reference point of a point represented by the recognition result is judged from a distance between both points, a threshold value of a distance that is decided according to the parameter is used for this criterion. Further, using a reference interval representing a restricted range on a map as a reference for judging a distance, and changing

and setting the reference interval including a reference point according to a prescribed parameter, a distance from the reference point representing the recognized result may be judged from the positional relation to the reference interval of the point (for instance, according to whether that point is within or without the reference interval). Here, an expanding/contraction magnification M of a map displayed on the screen may be used as the above-mentioned parameter and thus, it is advised to make the judging criterion (a distance threshold value) lower (smaller) when M becomes large and change and set the reference interval to a narrow area.

[0010]

Thus, when a distance judging criteria is changed and set according to a prescribed parameter, the distance can be judged according to, for instance, a designated scale (expansion/contraction magnification).

[0011]

Further, the present invention features that reliability of the recognized result that is subject to the distance judgment is judged and the distance is judged only when the recognized result is judged not reliable. Here, when reliability of the recognized result is judged with a normalized value based on a length of input voice corresponding to a collated score of the recognized result, it becomes possible to make a highly precise judgment.

[0012]

Thus, when the recognized result is judged as sufficiently reliable, it is possible to avoid to force a user to make an unnecessary checking operation by regarding it outside the generation of the confirming response irrespective of a distance (far or close) from or to the reference point.

[0013]

Further, this invention also features that the words representing points on the map for recognition are layered and controlled according to prescribed segments on the map and when the recognized result represents a point on the map, the layered segment that becomes a distance judging reference from that point is decided and the distance is judged according to that segment itself or the relation with a segment to which the reference point belongs (for instance, whether the reference point belongs to that segment).

Thus, when the judging reference is switched for each layer, it becomes possible to make the judgment of a distance without computing a distance.

[0014]

In addition, when an attribute designating a reference for the distance judgment is given to each recognition vocabulary representing a point on the map in advance and a distance of a point representing the recognized result

is judged according to the judging reference specified by the attribute of the recognized vocabulary, it becomes possible to change and set the judging reference for each vocabulary. In this case, it becomes also possible to judge a distance without computing it.

[0015]

Further, this invention relative to a voice recognition support method is also effected as an invention of a device (a voice recognition system).

Further, this invention is also effective as a computer readable recording medium recording a program for having a computer to execute procedures equivalent to the relevant invention (or a program for having a computer to function as a means equivalent to the relevant invention or a program for having a computer to realize a function equivalent to the relevant invention).

[0016]

[Preferred Embodiments of the Invention]

Preferred embodiments of this invention will be described below referring to the attached drawings.

[0017]

[First Embodiment]

FIG. 1 schematically shows a voice recognition system in a first embodiment of this invention.

Here, a voice recognition system is assumed. This system is applied to the voice recognition when making the

geographical searching by inputting a voice (a voice command) in a car navigation system. Vocabulary subject for voice recognition is restricted to place names and facility names in order to make the explanation simple and each vocabulary is given with a coordinate representing its geographic location (hereinafter, simply referred to as a position coordinate).

[0018]

The voice recognition system shown in FIG. 1 comprises a general controller 11, a reference point setting portion 12, a voice recognition portion 13, and a response generating portion 14. This response generating portion 14 has a distance judging portion 140.

[0019]

The general controller 11 controls the entire of the car navigation system including voice input from a voice inputting means such as a microphone (not shown), control of the display screen, setup and change of various parameters and further, control of required information databases.

[0020]

The reference point setting portion 12 sets up and maintains coordinate information (position coordinates) of place points (reference points) that become the bases for the geographical searching. For an initial value of the reference point, a position coordinate of a current

position of a car navigation or a pre-set prescribed place (an aimed place) is used. This position coordinate of the reference point can be changed and set according to a user's request or an autonomous operation of a system. Further, the position coordinate of the reference point is computed from the latitude and longitude obtained from the GPS (Global Positioning System) or the position coordinates registered in the database or a position on a map.

[0021]

The voice recognition portion 13 receives an input voice from a voice input means (here, a voice input means comprising a microphone and an A/D converter) that is controlled by the general controller 11, analyzes the input voice sound and obtains its characteristic pattern series, collates and computes the relevant characteristic pattern series with the standard characteristic pattern (the standard model) of recognized vocabulary, and outputs a recognition result. Here, when no recognized result could be obtained for the failure in voice detection, etc., the voice recognition portion 13 outputs the failure of recognition.

[0022]

A number of methods are already known as a definite collation method applicable in the voice recognition portion 13. Accordingly, any one of them can be selected and used. Definite examples of voice recognition are

available in detail in "Basis for Voice Recognition (Upper Vol.), (Lower Vol.)" written by L.Rabiner and Biign-Hwang Juang and edited by Tei Furui.

In this embodiment, in order to make the explanation simple, a case wherein the recognized result in the voice recognition portion 13 is only one and a place or a facility on a map is displayed will be described.

[0023]

The response generating portion 14 receives a recognized result (one) from the voice recognition portion 13 and produces a response to a user as shown below according to a flowchart shown in FIG. 2.

First, the distance judging portion 140 of the response generating portion 14 computes an Euclid distance D between the position coordinate at the reference point that is set and maintained by the reference point setting portion 12 and a position coordinate of a place name or a facility name (vocabulary) that is obtained as a recognized result (that is, the point shown by the recognized result) (Step S11).

[0024]

Then, the distance judging portion 140 of the response generating portion 14 compares the said distance D with a pre-determined threshold value T in order to judge a distance from the place point shown by the recognized result to the reference point based on the computed distance

between 2 points; that is, to make the judgment of a distance (far or near) from or to the reference point shown by the recognized result (Step S12).

[0025]

If D<T, the distance judging portion 140 judges that the point shown by the recognized result is near from the reference point and sets up a distance judging flag DF showing the judging result in the first state. In this case, the response generating portion 14 outputs the recognized result only to the general controller 11 without generating a response output that is described below according to the first state of the distance judging flag DF representing the distance judging result by the distance judging portion 140; in other words, according to the distance judging result showing that the point shown by the recognized result is near to the reference point (Step S13).

[0026]

On the other hand, if D is not larger than T, that is, $D \geqq T$, the distance judging portion 140 judges that the point shown by the recognized result is far from the reference point and sets up the distance judging flag DF in a second state. In this case, the response generating portion 14 generates a message to a user to confirm the recognized result (a confirming response), for instance "Is xx" (xx is a recognized result) according to the distance judging

result showing the point shown by the recognized result is far from the reference point and outputs a pair of this message (confirming message) and the recognized result to the general controller 11.

[0027]

When a pair of the confirming message and the confirmed result are notified from the response generating portion 14 (by Step S14), the general controller 11 presents the confirming message to a user through the voice output or the display screen and waits the confirmation (judgment) by a user. The user will input right or wrong of the recognized result through the button operation.

[0028]

When the user inputs that the recognized result is "in correct", the general controller 11 urges the user to speak again through the voice output or the display on the screen. Further, when "Correct" is input, the general controller 11 performs the operation corresponding to the recognized result.

[0029]

On the contrary, when a confirmation message was not contained in the response output from the response generating portion 14; that is, when the recognized result only is notified (in Step S13) or the failure in the recognition is notified, the general controller 11 makes the operation corresponding to the notified contents

without waiting the confirmation by the user.

[0030]

In the embodiment described above, it is limited to a case wherein a voice recognition vocabulary has position coordinates in order to simplify the explanation. However, voice recognition vocabulary generally contains those words having no position coordinate such as a system control command name, etc.

[0031]

So, in the voice recognition system shown in FIG. 1, to enable to correspond even when the recognized result has no position coordinate in the vocabulary, it is better to make the system in such a structure that a position coordinate flag PF to show whether a word has a position coordinate is given to all of the words as an attribute and the recognized result attached with the position coordinate flag PF is given to the response generating portion 14 from the voice recognition portion 13. The operation of the response generating portion 14 in this structure will be described below.

[0032]

First, the distance judging portion 140 of the response generating portion 14 examines the position coordinate flag PF attached to the recognized result given from the voice recognition portion 13 and judges if the recognized result is a word having a position information.

[0033]

If a vocabulary is a word having no position information (the recognized result), the distance judging portion 140 regards the distance, for example, as $D \geqq T$ (definitely, $D=T$) and sets the distance judging flag DF in the second state. In this case, the confirming message paired with the recognized result are output from the response generating portion 14 to the general controller 11 as clear from the flowchart shown in FIG. 2. On the contrary to the above, regarding the distance to be $D<T$ (definitely, $D=0$), the distance judging flag DF may be set in the first state. In this case, the recognized result only is output to the general controller 11 from the response generating portion 14. In addition, a distance may be decided to $D \geqq T$ ($D=T$) or $D<T$ ($D=0$) according to the inner state of a car navigation system.

[0034]

[Second Embodiment]

Next, a voice recognition system in a second embodiment of this invention will be explained.

The voice recognition system in this second embodiment is characterized in that the voice recognition portion is capable of outputting plural recognition candidates as the voice recognized result when compared with the voice recognition portion in the first embodiment, which outputs only one voice recognized result and that the response

generating portion and the general controller are provided with new functions corresponding to the output function of plural recognized candidates. However, the construction shown in FIG. 1 will be employed for the convenience.

[0035]

The operation of the voice recognition system in the second embodiment will be described below centering on the processing in the response generating portion 14 and the general controller 11.

First, the processing of the response generating portion 14 referring to a flowchart in FIG. 3.

[0036]

Now, it is assumed that plural recognition candidates are given to the response generating portion 14 from the voice recognition portion 13 as recognized results. Further, it is also assumed that these recognition candidates are attached with a collation score S as an evaluation value (a pattern matching result representing a similarity or a distance to an input voice (its characteristic pattern series) and the said candidate (its standard characteristic pattern) and a position information flag PF showing whether the said candidate is a word having position information.

[0037]

The distance judging portion 140 of the response

generating portion 14 judges first the first ranked recognized candidate as to whether it has a position coordinate from the attached position information flag PF (Step S21).

When the first ranked candidate is a word having no position coordinate, the response generating portion 14 outputs the first ranked candidate only to the general controller 11 as a recognized result (Step S22) and processes the recognized result in the same manner as in the first embodiment.

[0038]

On the other hand, when the first ranked candidate has a position coordinate, the distance judging portion 140 of the response generating portion 14 extracts only candidates having position coordinates from the recognized candidates given from the voice recognition portion 13 (Step S23).

[0039]

In succession, the distance judging portion 140 obtains a distance D between a reference point (that is set and maintained by the reference point setting portion 12) and the said candidates for each of extracted candidates and obtains a new score S' according to the following equation (1) from the distance D and the said candidate's score S,

$$S' = \alpha S + \beta G(D) \quad \ldots (1)$$

then, based on this score S', arranges the candidates in

the order of larger score (Step S24). Further, $\alpha$ and $\beta$ are experimentally decided coefficients, G (D) is a function for D and a simple decreasing function or simple non-increasing function.

[0040]

In the re-arrangement of the candidates in the above Step S24, when, for instance, $\alpha=0$, $\beta>0$ and G (D) is a simple decreasing function of D in the above equation (1), it is synonymous to rearrange candidates in order of close to the reference point irrespective of the candidate score S. Further, when $\alpha>0$ and $\beta>0$, it is synonymous to no score computation. Further, it is not always necessary to use Equation (1) to obtain the score S' and Score S' may be computed from at least a distance D and a score S.

[0041]

When the candidate ranking is rearranged by the re-scoring in the above Step S24, the distance judging portion 140 computes a distance D between the point shown by the candidate that becomes the first rank by the rearrangement and the reference point (Step S25). Then, the distance judging portion 140 judges a distance far or near from the point shown by the first rank candidate by comparing the computed distance D with a threshold value, and sets the distance judging flag DF in the first state (when $D<T$) or the second state ($D \geqq T$) according to the result of the judgment (Step S26).

[0042]

When the distance judging result by the distance judging portion 140 is D<T, the response generating portion 14 notifies the recognized result of the first ranked candidate only to the general controller 11 (Step S22). On the contrary, when the distance judging result is D ≧T, the response generating portion 14 notifies a pair of the recognized results corresponding to t confirmation message similar to that in the first embodiment for each of N pieces of candidates to the upper N ranking candidate (N is a pre-determined natural number) (when the number of candidates is less than N, all candidates) to the general controller 11 (Step S27). Further, when no recognized result was obtained by the voice recognition portion 13 and the failed recognition is notified to the response generating portion 14 by the voice recognition portion 13, the failed recognition is also notified to the general controller 11 from the response generating portion 14.

[0043]

The general controller 11 executes the process corresponding to the notified contents from the response generating portion 14 as shown below according to a flowchart in FIG. 4.

[0044]

First, the general controller 11 judges whether the recognition is failed based on the notified contents from

the response generating portion (Step S31). If the recognition was failed, the general controller 11 executes the process corresponding to the failed recognition (Step S32). Further, the process itself when the recognition was failed is not related directly to this invention and its explanation is omitted.

[0045]

On the contrary, when the recognition was not failed, the general controller 11 sets an order number n representing the ranking of the recognition candidate at an initial value 1 (Step S33). When a value of N is not over the number of candidates as in this example (Step S35), the general controller 11 checks if a confirmation message was attached to the N-th rank candidate (Step S36). If no confirming message was attached to the n-th rank candidate, the general controller 11 executes the same process as in the recognition failure (Step S32). Such the candidate having no confirmation message exists only when the recognized result of the first rank candidates only was notified from the response generating portion 14 (Step S22).

[0046]

On the other hand, when a confirmation message was attached to the n-th ranked candidate, the general controller 11 presents the said confirming message to a user by the voice output or the screen display and waits

the confirmation (judgement) of the user (Step S37). The candidates having such confirming messages exit only when the candidates up to the higher N-th rank paired with confirming messages are notified from the response generating portion 14 (in the above Step S27).

[0047]

When a user inputs "Correct" of the n-th rank candidate in response to the confirming message presented to the user, the general controller 11 executes the operation corresponding to the n-th ranked candidate (the recognized result (Steps S38, S39).

[0048]

On the contrary, when a user inputs "Incorrect" of the n-th candidate, the general controller 11 increases N by one (Steps S38, S34) and checks whether a value of N after the increment exceeds the number of candidates (Step S35). If the value of n after the increment is not above the number of candidates, the general controller 11 executes the processes subsequent to Step S36 for the n-th ranked candidate having a confirming message shown by an n value.

[0049]

Thus, when candidates to the higher N-th rank (all candidates when the number of candidates is less than N) paired with a confirming messages, respectively are notified from the response generating portion 14, the general controller 11 repeats the operation to present a

confirming message to a user orderly from higher rank candidates until a user input "Correct". When a user inputs the " correct" , regarding the n-th candidate at that time as the recognized result, the general controller 11 executes the operation corresponding to the recognized result. Further, when a user does not consent all candidates (Step S35), the general controller 11 executes the same processes as in the recognition failure (Step S32).

[0050]

[Third Embodiment]

Next, a voice recognition system in a third embodiment of this invention will be explained.

In a car navigation system, a scale of a map displayed on the screen is generally variable. In this type of system, the scale of a map is changed so as to display a designated area by expanding or contracting by the direction of a user or the control of a car navigation system. When making the voice input by a user, it is expected that a user speaks a name of place (or a facility) in a designated range in the most cases. In this case, when a coordinate of a reference point was not changed even if the display scale of a map was changed, it becomes difficult to judge a distance (far or near) correctly.

[0051]

The voice recognition system involved in the third embodiment features that the correct distance can be judged

when the map display scale was changed. In this case, the function of the reference point setting portion and that of the response generating portion partially differ from the function of the voice recognition system involved in the first embodiment. However, the construction shown in FIG. 1 is quoted for convenience.

[0052]

The operation of the voice recognition system involved in the third embodiment will be described below centering around the reference point setup by the reference point setting portion 12 and the distance judgment by the response generating portion 14. A case to apply it to a car navigation system capable of the map display by the variable scale is taken as an example.

First, it is assumed that the map scale was changed to display a designated area by expanding or contracting by the user's direction or the control of a car navigation system. In this case, the reference point setting portion 12 sets the central point of a map displayed by expanding or contracting as the reference point.

[0053]

Then, the distance judging portion 140 of the response generating portion 14 acquires a map expansion/contracting magnification M (the more M is large, the more narrow the area is limited) and sets a threshold value T as shown by the equation (2) based on the value of M. Here, F(M) is

a simple non-increasing function of M.

$$T = F(M) \quad \ldots (2)$$

Here, F(M)is a simple non-increasing function of M.

[0054]

Now, it is assumed that a user speaks a name of a desired place (or a facility) on a map and the recognized result of the spoken contents by the voice recognition portion 13 is given to the response generating portion 14. Here, to make the explanation simple, it is assumed that one voice recognized result only was given.

[0055]

The distance judging portion 140 of the response generating portion 14 computes a distance D between the point shown in the recognized result given from the voice recognition portion 13 and the reference point that was set by the reference point setting portion 12, and judges a distance to the point shown by the recognized result by comparing the computed D with the previously set threshold value T. As clearly known, if a distance is D>T, it is judged far from the reference point.

[0056]

A case to change and set a threshold value based on the expansion/contraction magnification M in a car navigation system with variable expansion and contraction map display magnification is explained above but the system is not restricted to this. For example, by introducing

a concept of the reference range for judging a distance (an unconditional searching subject area generating no confirming response) and the said reference range is changed and set according to the expanding/contracting magnification M, a point shown in the recognized result may be judged if far or near from the reference point according to whether the point is in the said reference range. Here, it may be better to set the reference range centering on the reference point. Further, prescribed parameters representing a reference for judging a distance (threshold value, reference range) may be introduced for the expanding/contracting magnification M.

[0057]

[Fourth Embodiment]

In the embodiments described above, a distance D between a reference point and a point shown in the recognized result (hereinafter, the recognized result represents the first candidate after the re-scoring when there are plural candidates as in the second embodiment) was obtained by a Euclid distance from a coordinate. However, this distance D is used for judging a distance between two points and therefore, not limited to the Euclid distance and can be a scale indicating a distance between two points.

[0058]

So, a voice recognition system in a fourth embodiment of this invention that uses a scale indicating a distance

between two points instead of Euclid distance as a distance D between a reference point and a point shown in the recognized result (hereinafter, called as a distance between a reference point and a recognized result will be explained below by employing the construction shown in FIG. 1 for convenience taking the state in Japan as an example.

[0059]

First, in this embodiment, words that are subject to recognition and have position coordinates are layered in order from higher level addresses (prefecture, city, town, village, block number, facility) and each word is able to show each of addresses to which it belongs. Further, when a position coordinate only can be obtained from the information sent from the GPS, for example, a current location, an address most near to a position coordinate in Euclid distance is obtained. The current position belongs to that address.

[0060]

In such the example, the distance judging portion 140 of the response generating portion 14 regards a distance D between the recognized result of the voice recognition portion 13 and the reference point as D=0 when both belong to the same prefecture and D = T when both do not belong to the same prefecture (T is a threshold value for the distance judgement).

[0061]

Therefore, in the response generating portion 14, when both the reference point and the recognized result are not in the same prefecture, D is regarded as T (D=T) and a confirming message is generated as clear, for example, from Steps S12 and S14 in FIG. 2. On the contrary, when the reference point and the recognized result are in the same prefecture, D is regarded as 0 (D=0) and therefore, in the response generating portion 14. The recognized result only is output as clear in Steps S12 and S13 shown in FIG. 1.

[0062]

When the distance D computing (deciding) technique described above is used, the distance judgment by the distance judging portion 140 of the response generating portion; that is, the comparison of the distance D and the threshold value T is synonymous with the judgment that is made depending on whether the distance D and the threshold value are in the same prefecture without computing D. Therefore, the distance judging portion 140 may judge a distance depending on whether the reference point and the recognized result are in the same prefecture without computing D instead of judging a distance of place or facility of the recognized result by computing (deciding) a distance D between the reference point and the recognized result and comparing the distance D with the threshold value T according to the above-mentioned technique.

[0063]

Further, the computation of a distance and the judgment of a distance are made here according to whether the reference point and the recognized result are in the same prefecture but they may be in a city, town or village instead of prefecture. However, when a city, town or village is used for the distance judging standard, it is necessary to consider that if the recognized result is a prefecture name, it becomes impossible to judge whether the recognized result is in the same city, town or village as the reference point; that is, when the recognized result is a word belonging to the higher layer than the judging standard, a distance cannot be judged. So, in such the case, the reference point and the recognized result are treated as not being in an area represented by a word of the same layer and a confirming message may be generated in the response generating portion 14 regarding D=T, that is, a place or a facility shown by the recognized result is far from the reference point.

[0064]

Further, a local layer higher than prefecture, for example, the Kanto district or the Tokai district or a layer differing from generally known geographical classification or administrative unit may be provided and the computation or judgment of distances can be made.

[0065]

[Deformed Embodiment of Fourth Embodiment]

In the fourth embodiment described above, a distance judging standard is pre-determined, for instance, prefecture, etc. However, the judging standard is not restricted to this but different standards may be used according to the recognized result.

[0066]

So, a deformed embodiment of the fourth embodiment to switch and set distance judging standards according to a recognized result in the case of Japan will be explained employing the construction shown in FIG. 1 for convenience.

[0067]

Here, it is assumed that there is no higher layer than a prefecture name. The distance judging portion 140 of the response generating portion 14 regards D=T unconditionally when the recognized result by the voice recognition portion 13 is a prefecture name; that is, the recognized result of a place or facility is far from the reference point and has the response generating portion 14 generate a confirming message. Further, when the recognized result is a city, town or village, the distance judging portion 140 judges a distance according to whether the reference point and a city, town or village of the recognized result are in the same prefecture. That is, the distance judging portion 140 judges a distance at a layer above the recognized result by one according to whether it agrees the reference point. Further, when the

recognized result is at the highest layer, a distance may be regarded as D=T or D=0 unconditionally as in the case of prefecture described above.

[0068]

In addition, instead of the layer unit, the judging standard may be switched for each recognizing word or category, for example, 『○○ Gas Station XX Town Shop』 as the city, town, village level or 『□□ Amusement Park』 as the prefecture level. Furthermore, a distance may be regarded D=T or D=0 unconditionally depending on a word. Such switching of judging standards can be achieved, for instance, when an attribute for designating a judging standard is given to a recognized word.

[0069]

[Fifth Embodiment]

In the embodiments described above, the recognized result having a position coordinate is always subject to the distance judgment by the distance judging portion 140 of the response generating portion 14 except the judging result is decided unconditionally. However, when the recognized result is sufficiently reliable, it is possible to avoid to force an unnecessary confirming operation when a confirming response to a user is not generated independently of a distance from the reference point.

[0070]

So, a voice recognition system involved in a fifth

embodiment of this invention to control the confirming response generation based on the reliability of the recognized result will be explained using the construction shown in FIG. 1 for convenience and referring to a flowchart shown in FIG. 5.

[0071]

Now, it is assumed that plural recognition candidates are given to the response generating portion 14 as recognized results from the voice recognition portion 13. It is further assumed that these recognition candidates are provided with a collated score S that is an evaluation value representing similarity of characteristic pattern series of input voice with the standard characteristic patterns of said candidates or distance between them, a length (time) T of said voice and a position information flag PF showing whether said candidates are words having position information.

[0072]

The distance judging portion 140 of the response generating portion 14 first judges a first ranked candidate as to whether it is a word having a position coordinate from the position information flag PF (Step S41). The operation of the first ranked candidate that is a word having no position coordinate is the same as, for instance, Step S22 in FIG. 3 and the first ranked candidate only is output to the general controller 11 from the response generating

portion 14 (Step S42).

[0073]

On the other hand, when the first ranked candidate has a position coordinate, the distance judging portion 14 of the response generating portion 14 computes reliability R of the said first ranked candidate differing from the second embodiment (refer to the flowchart in FIG. 3). This reliability R will be described below.

[0074]

First, the collated score S that is the recognized result acquire in the voice recognition portion 14 is an accumulated value of scores for each unit time (for example, a frame period). Accordingly, reliability of the recognized result is not determined only from a size of the collation score S. So, in this embodiment, reliability R of the recognized result (the firs rank candidate) is obtained from, for instance, the following equation:

$$R-S/T \quad \ldots (3)$$

where T is a length (time) of voice that becomes the subject for recognition and is given with the score S from the voice recognition portion 13 as described above. Reliability R according to the above equation (3) is made by the distance judging portion 140 but may be made by the voice recognition portion 13.

[0075]

When reliability R of the first ranked candidate is

computed, the distance judging portion 140 compared the said reliability R compares with a pre-determined threshold value A and judges whether the first ranked candidate is reliable or not (Step S44).

[0076]

If R is larger than A (R>A), the distance judging portion 140 judges that the first ranked candidate (the recognized result) is sufficiently reliable. In this case, the response generating portion 14 does not generate a confirming response likewise when a recognized result has no position coordinate or a distance between the recognized result and the reference point is smaller than T, and outputs the first ranked candidate only to the general controller 11 (Step SS42).

[0077]

On the other hand, when R is smaller than A ($R \leqq A$), the distance judging portion 140 judges that the first ranked candidate (the recognized result) is not reliable. In this case, the distance judging portion 140 executes the distance judgment according to the processing steps S45 ~ S48 similar to Steps S23 ~ S26 in FIG. 3 in the second embodiment, and judges the confirming response generation.

[0078]

The first through the fifth embodiments of this invention are explained in the above. However, the functions of the general controller 11, the reference point

setting portion 12, the voice recognition portion 13 and the response generating portion 14 can be realized in a software.

[0079]

Further, this invention can be applied as a recording medium for computer readable CD-ROM, etc. recording programs for having a computer to execute procedures applied in the voice recognition system involved in the above-mentioned embodiments; especially prescribed procedures including a distance judging and confirming message (confirming response) generating process and a confirming message presenting process to a user in the general controller 11 (or having a computer to function as a prescribed means retained by a voice recognition system or having a computer to realize prescribed functions retained by the voice recognition system). Further, this program may be that can be downloaded through a communication medium.

[0080]

In addition, the embodiments of this invention can be modified variously from the examples described above. These modifications are within the range of the embodiments of this invention unless departing from the spirit and scope thereof.

[0081]

[Effects of Invention]

As described above in detail, according to this invention, when it is judged that a point shown in the recognized result to the input voice, a confirming response is presented to a user, urging the confirmation of the recognized result and requesting a user's direction. Therefore, it is possible to avoid unnecessary erroneous operations for erroneous recognition and improve the usability of the voice recognition system without limiting places on a map that can be vocally designated; that is, without limiting an area subject to search.

[Brief Description of Drawings]

[FIG. 1]

A block diagram showing the schematic construction of a voice recognition system involved in one embodiment of this invention.

[FIG. 2]

A flowchart for explaining the processing procedures of a response generating portion 14 when the voice recognition portion 13 in FIG. 1 outputs only one recognized result.

[FIG. 3]

A flowchart for explaining the processing procedures of the response generating portion 14 when the voice recognition portion 13 in FIG. 1 outputs plural candidates as the recognized result.

[FIG. 4]

A flowchart for explaining the processing procedures by the general controller 11 in FIG. 1 corresponding to the notified contents from the response generating portion 14 that operates according to the flowchart shown in FIG. 3).

[FIG. 5]

A flowchart for explaining the deformed example of the processing procedures of the response generating portion 14 when plural candidates are output as the recognized result by the voice recognition portion 13 in FIG. 1.

[Description of Reference Numbers]

11 ... General controller (Presenting means)

12 ... Reference point setting portion

13 ... Voice recognition portion

14 ... Response generating portionm

14C .. Distance judging portion (Re-scoring means)

[Document] Abstract

[Abstract]

[Object] To avoid unnecessary erroneous operations for erroneous recognition and improve usability of the system without limiting an area subject to search that can be vocally designated.

[Construction] A distance between a point shown by he voice recognized result (for example, place or facility name) for a voice spoken by a user for searching on a map and a reference point that is set prior to the recognition is computed (Step S11), a distance from a point shown by the recognized result to the reference point is judged by comparing sizes of the distance D and a threshold value T (Step S12), and when the point is judged far from the reference point ($D \geqq T$), a confirming message urging a user to confirm right or wrong of the recognized result a user (Step S14) and the message is presented to a user.

[Selected Drawing] FIG. 2

[Document] Drawings

[FIG. 1]

11 General controller

12 Reference Point Setting Portion

13 Voice Recognizing Portion

14 Response Generating Portion

140 Distance Judging Portion

[FIG. 2]

Recognized Result (Place Name, Facility Name)

S11 Compute a distance D between recognized result and reference point

S12 D<T (threshold value) ?

Yes

S13 Output recognized result

No

S14 Output a confirming message and a recognized result

End

[FIG. 3]

Candidate

S21 1st rank candidate has a coordinate ?

No Yes

S23 Extract only a candidate having a coordinate

S24 Re-score extracted candidates and change ranks by new scores

S25 Compute a distance D between 1st rank candidate and rerence point

S26 D<T (threshold value) ? Yes

S22 Output 1st rank candidate No

S27 Output confirming response and candidates up to higher N rank

End

[FIG. 4]

Response generating portion output (candidate, recognition fail)

S31 Recognition failed ? Yes

No

S35

n>Number of candidates ? Yes

S36 Is there a confirming message for n-th rank candidate ? No

Yes

S37 Output a confirming message and wait user's judgment

No

Yes

S38 n-th rank candidate is correct ?

No

Yes

S39 Execute process corresponding to recognized result

S32 Execute process corresponding to failed recognition

End

[FIG. 5]

Candidate

S41 1st rank candidate has a coordinate ? No

S42 Output 1st rank candidate

S43 Compute reliability R

S44 R>Threshold value ? Yes

No

S45 Extract only candidates having a coordinate

S46 Re-score extracted candidates and change ranks by new scoores

S47 Compute a distance D between 1st rank candidate and reference point

S48 D<T (Threshold value) ? Yes

No

S49 Output confirming response and candidates up to higher N rank

End